**Objective:** To demonstrate that a consensus approach for combining prediction equations based on clinical and exercise test variables derived from different populations can stratify patients referred for possible coronary artery disease (CAD) into low-, intermediate-, and high-risk groups.

**Design:** Retrospective analysis of consecutive patients with complete data from exercise testing and coronary angiography referred for evaluation of possible CAD. After derivation of a logistic equation in our own training set of patients, this equation, along with two other equations developed independently by other investigators, was validated in a test set. The validation strategy for the consensus approach included the following: (1) calculation of probability scores for each patient using each logistic equation independently; (2) determination of probability thresholds in the training set to divide the patients into three groups—low risk (prevalence CAD <5%), intermediate risk (5 to 70%), and high risk (>70% prevalence of CAD); (3) using agreement among at least two of the three prediction equations to generate “consensus” for each patient; and (4) application of the consensus approach thresholds to the test set of patients.

**Settings:** Two university-affiliated Veteran’s Affairs medical centers.

**Patients:** We studied 718 consecutive men between 1985 and 1995 who had coronary angiography within 3 months of an exercise treadmill test for suspected CAD. The population was randomly divided into a training set of 429 patients and a test set of 289 patients. Patients with previous myocardial infarction or coronary artery bypass surgery, valvular heart disease, left bundle branch block, or any Q waves present on their resting ECG were excluded from the study.

**Measurements:** Recording of clinical and exercise test data along with visual interpretation of the ECG recordings on standardized forms and abstraction of visually interpreted angiographic data from clinical catheterization reports.

**Results:** We demonstrated that by using simple clinical and exercise test variables, we could improve on the standard use of ECG criteria during exercise testing for diagnosing CAD. Using the consensus approach divided the test set into populations with low, intermediate, and high risk for CAD. Since the patients in the intermediate group would be sent for further testing and would eventually be correctly classified, the sensitivity of the consensus approach is 94% and the specificity is 92%. The consensus approach controls for varying disease prevalence, missing data, inconsistency in variable definition, and varying angiographic criterion for stenosis severity. The percent of correct diagnoses increased from the 67% for standard exercise ECG analysis and from the 80% for multivariable predictive equations alone to >90% correct diagnoses for the consensus approach.

**Conclusions:** The consensus approach has made population-specific logistic regression equations portable to other populations. Excellent diagnostic characteristics can be obtained using simple data and measurements. The consensus approach is best applied utilizing a programmable calculator or a computer program to simplify the process of calculating the probability of CAD using the three equations.

*(CHEST 1997; 111:1742-49)*

**Key words:** angiography; coronary disease; exercise testing; prediction equations

**Abbreviations:** CAD=coronary artery disease; LB-PA=Long Beach and Palo Alto; MET=metabolic equivalent; ROC=receiver operating characteristic; VA=Veterans Affairs

*From the Cardiology Division (Dr. Do and Drs. West, Atwood, and Froelicher), Veterans Affairs Palo Alto Health Care System, Stanford University, Palo Alto, Calif, and West Virginia University Medical Center (Dr. Morise), Morgantown.*

Manuscript received November 1, 1996; accepted November 4.

Reprint requests: Victor Froelicher, MD, Cardiology Division, (111C), VA Palo Alto Health Care System, 3801 Miranda Ave, Palo Alto, CA 94304; email: vicemd@aol.com
With the emergence of managed care and capitation, the validation of clinical strategies using common diagnostic modalities like exercise testing to direct patients with possible coronary disease to appropriate levels of evaluation and treatment becomes essential. The application of multivariate analysis using discriminate function and logistic regression techniques to clinical and exercise test variables has been shown repeatedly to improve on the standard application of the exercise ECG test to diagnosing coronary artery disease (CAD). Statistical techniques that combine the patient’s medical history, symptoms of chest pain, hemodynamic data, exercise capacity and exercise ECG response have been proved to be better predictors for CAD than a single ECG criterion like ST segment depression. Despite the validation of prediction equations in large patient samples, however, the method has not been widely disseminated.

Clinicians remain skeptical regarding the applicability of prediction equations to clinical practice. The variability in disease prevalence among populations with suspected CAD, the lack of standards for defining and capturing clinical data, and the unavailability of an efficient mechanism for calculation of scores make radioisotope imaging or echocardiography preferable. Although the widespread use of computers for patient data management may remove the impediment of probability score calculation, concerns about disease prevalence and severity, definition of discriminant variables, missing data, angiographic and exercise testing methods, and other factors could affect the portability of these equations to other populations and thus limit their applicability for clinical practice.

Although the accuracy of prediction equations has been tested by applying receiver operating characteristic (ROC) techniques to the equations in other populations, calibration remains an issue. For example, one equation may predict an 80% probability for disease in a specific patient while another equation predicts a 50% probability even though both equations can discriminate equally between those with and without CAD. To enhance calibration, investigators have suggested that calibration be corrected by the disease prevalence in the clinical population in which the equation is applied. This may be difficult, however, since most clinicians do not know the disease prevalence in their exercise laboratory and even if they did, it could change from month to month.

A hypothetical way to make these equations more portable and self-calibrating would be to require a consensus between a number of equations for patient classification as low or high risk. The consensus approach has been used by NASA to calculate the trajectories of spacecraft but to our knowledge has not been applied in health care. Patients would be classified only as low or high risk if at least two of three equations agreed; i.e., consensus by majority. To explore this method, we performed this study in three steps. First, we derived equations predicting CAD before and after treadmill testing in a training set. Second, we applied our derived prediction equation along with two other equations derived and validated in large patient samples by Morise et al and Detrano et al to classify patients into high, intermediate, or low risk of CAD by consensus. Third, we demonstrated that this approach performed successfully in a test set of patients and that it functioned with varying prevalences of disease and variations in angiographic criteria.

**Materials and Methods**

**Clinical Assessment**

From a target population of 4,099 patients, we selected 718 consecutive male patients who underwent exercise testing for suspected coronary disease at the Long Beach Veterans Affairs (VA) or the Palo Alto VA Medical Centers between 1985 and 1995 who fulfilled the following criteria: (1) inclusion criteria—complete data with coronary angiography results obtained within 3 months of the exercise test; and (2) exclusion criteria—previous myocardial infarction or coronary artery bypass surgery, valvular heart disease, left bundle branch block, or any Q waves present on the resting ECG.

The population was randomly divided into a training set of 289 patients and a test set of 289 patients. The clinical variables obtained from the initial history included age, symptoms of chest pain, body mass index, obesity, current cigarette smoking and number of pack-years of cigarette smoking, family history of CAD, and history of congestive heart failure, hypertension, noninsulin- or insulin-dependent diabetes mellitus, stroke, peripheral vascular disease, hypercholesterolemia, and COPD. Chest pain symptoms were coded as 1 for typical, 2 for atypical, 3 for nonanginal pain, and 4 for no chest pain. All other clinical variables except for age, body mass index, and pack-years of cigarette smoking were coded as 1 for presence and 0 for absence.

**Exercise Testing**

All patients underwent exercise testing using an individualized ramp or the United States Air Force School of Aerospace Medicine protocol. Before ramp testing, we assessed a patient’s exercise capacity with a questionnaire listing activities in an increasing order according to expected metabolic expenditure. Using this specific activity questionnaire to adjust the exercise protocol allowed most patients to reach maximal exercise at approximately 10 min. ST segment depression was measured at the J junction and ST slope measured over the following 60 ms was classified as upsloping, horizontal, or downsloping. A code of 1 was used for abnormal slope (horizontal or downsloping and ST depression of at least 0.5 mm) and 0 for normal slope (upsloping or ST depression of <0.5 mm). In addition, all of the following hemodynamic measurements were recorded: resting and maximal heart rate, change in heart rate (maximal heart rate—rest
heart rate), resting and maximal systolic BP, change in systolic BP, maximal double product (maximal heart rate × systolic BP/1,000), exercise-induced hypotension (a drop in exercise systolic BP below standing or a drop in systolic BP of 20 mm Hg after a rise), exercise-induced angina, and exercise capacity estimated in metabolic equivalents (METs) from the final treadmill speed and grade. Angina during testing was classified according to the Duke Exercise Angina Index (2 if angina required stopping the test, 1 if angina occurred during or after treadmill test, and 0 for no angina). 15 No test result was classified as indeterminate; 16 treatment with medications was not withheld, and no heart rate targets were applied. 17

Coronary Angiography

Coronary artery narrowing was visually estimated and expressed as percent luminal diameter stenosis. Patients with a 70% narrowing in the left anterior descending, left circumflex, or right coronary arteries or their major branches, or a 50% narrowing in the left main coronary artery were classified as having significant angiographic coronary artery disease. We used the 70% criterion because of its pathophysiologic association with coronary flow limitations during stress 18 and its widespread use in other clinical trials like the Coronary Artery Surgery Study. 19

Statistical Analysis

Patients in the Long Beach and Palo Alto VA (LB-PA) training set were divided into two groups, with and without coronary artery disease. Initially, the clinical and resting ECG variables in the two groups were compared using an unpaired t test. The variables with a p value <0.05 were selected into a stepwise logistic regression model for predicting the probability of CAD. Software (True Epistat Software; Richardson, Tex) was used for the logistic regression procedures. To take the shape of a sigmoid curve, representing the relation of the dose response of most risk factors to atherosclerosis formation, the model consisted of the following natural log equation yielding a number from 0 to 1 for 0 to 100% probability for CAD:

\[
\text{Probability} = \frac{1}{\left(1 + e^{-\left(a + bx + cy + \ldots \right)}\right)}
\]

where a is the intercept, b and c are coefficients, x and y are the values for the variables found to be significantly associated with disease.

After developing a model for predicting preexercise test probability for CAD, the exercise test variables were considered along with the preexercise score as candidates for a second stepwise logistic regression to predict postexercise test probability for CAD, resulting in the LB-PA logistic regression equation.

Consensus Approach

After development of the LB-PA logistic regression equation, we calculated a probability score for each patient in the training set with the LB-PA equation and two other validated equations developed in different populations from Morise et al 10 and Detrano et al, 11 respectively, using each equation independently for each patient. For details of the calculation of the equations of Morise et al and Detrano et al, please see the Appendix. We then partitioned the population repetitively using probability scores from 20 to 90 in increments of 5 for each equation separately. Using a consensus strategy that required at least two of the three logistic equations to agree at each probability score level, we separated the training set into groups using all equations at each probability score, calculating the prevalence of CAD in each group below or above the chosen probability score. We chose two thresholds or probability scores from results in the training set that separated the population into three groups: low risk (prevalence CAD <5%), intermediate risk (5 to 70%), and high risk (>70% prevalence for CAD). Finally, these thresholds were applied to the test set using the consensus approach and test performance characteristics assessed.

Results

Derivation of the LB-PA Logistic Regression Equation

Preexercise Test Equation: The clinical variables and resting ST depression in patients with and without CAD in the training set were compared. Significant differences in patients with CAD included an older age, more cigarette consumption, and a significantly higher prevalence of typical angina, hypercholesterolemia, and diabetes mellitus. All variables with p values <0.05 in the training set were entered into the logistic regression model. The preexercise test equation including the chosen variables and their coefficients, and the constant is given below:

\[-2.1 + (0.03 \times \text{age}) -(0.4 \times \text{symptoms}) + (0.8 \times \text{diabetes}) + (0.4 \times \text{hypercholesterolemia}) + (0.01 \times \text{pack-years}) + (0.7 \times \text{resting ST depression in mm})\]

Postexercise Test Equation: The exercise test data between patients with and without angiocardiographic coronary disease in the training set were then compared. The group with CAD had more ST depression and a lower maximal heart rate and exercise capacity. Although most of the exercise variables were good univariate predictors, only four variables were independent predictors when considered in the multiple logistic model: exercise ST depression, ST slope, exercise capacity in METs, and the presence of angina during exercise. The postexercise test equation, including the chosen variables, their coefficients, and the constant, is given below:

\[-1.2 + (3.3 \times \text{pretest}) + (0.5 \times \text{exercise ST depression in mm}) + (0.6 \times \text{ST slope}) - (0.16 \times \text{METs}) - (0.5 \times \text{exercise angina})\]

(Pretest is a number between 0 and 1 generated by the pretest equation.)

Validation of the LB-PA Equation

We validated the new LB-PA equation by comparing test performance characteristics using ROC curves among the training and test sets and comparing ROC curves among the new LB-PA equation and...
two other validated logistic equations from Morise et al\textsuperscript{10} and Detrano et al.\textsuperscript{11} The area under the ROC curves represents the discriminatory power of the equation.

Comparison of Training and Test Groups: There were no significant differences between training and test groups when clinical and exercise parameters were compared.

ROC Area Analysis: ROC curves were generated from the probability scores derived from the preexercise and postexercise equations. The ROC curve area for the preexercise equation was 71% (±2%). The area increased to 77% for the postexercise equation, which included the exercise test variables. The multivariate analysis method significantly improved the prediction of CAD over the standard ST criteria of I mm horizontal or downsloping ST depression at the J junction, which had an area of 67% (p<0.0001). Along with the probability score generated from the Palo Alto equation, we also calculated the probability scores using the equations of Detrano et al\textsuperscript{11} and Morise et al\textsuperscript{10} for each patient in the training set. These equations had an area of 76%, similar to our newly derived LB-PA equation.

Consensus or Classification by Agreement of the Equations From LB-PA, Detrano et al, and Morise et al

The patients were classified in the training set by probability score using the equations from LB-PA, Detrano et al.\textsuperscript{11} and Morise et al.\textsuperscript{10} independently and in consensus (agreement of at least two of three equations), comparing the prevalence of CAD among groups at each probability score (Table 1). For example, 137 patients had a LB-PA probability score of ≤30%, and 16% of these 137 patients had CAD. In comparison, 19 and 85 patients had the Detrano et al and Morise et al probability scores of ≤30%, respectively. Similarly, using the consensus approach, a total of 70 patients had probability scores of ≤30% in two of these three scores (LB-PA, Detrano et al, and Morise et al); 4% of this group had CAD.

Using the results in Table 1, threshold values were determined based on probability scores that partitioned the population into groups by prevalence of CAD: <5% (low risk); 5 to 70% (intermediate risk); >70% (high risk). Patients with a probability ≤30% were classified as low risk for CAD; patients with a probability ≥75% were classified as high risk for CAD. The remaining patients were placed in the intermediate-risk group for CAD (see Table 2). For example, 126 patients or 29% of the training set would be classified as high risk; the calculated disease prevalence in this group is 71%.

Validation of Classifying the Patients by the Consensus Approach

The consensus classification was applied to the test set and the results are listed in Table 3. This comparison confirmed the performance in the training set.

Comparing the Consensus Approach With Standard ST Criteria

The predictive accuracy in the test set using standard ST criteria was 67%. Using the consensus classification by the three equations, 56, 149, and 84 patients were placed in the low-, intermediate-, and high-risk groups for CAD, respectively. The consensus approach misclassified 12 patients as false-positive and eight patients as false-negative in the test set. If the 149 patients in the intermediate risk group

<table>
<thead>
<tr>
<th>Probability by Scores</th>
<th>LB-PA Equation</th>
<th>Equation of Detrano et al\textsuperscript{11}</th>
<th>Equation of Morise et al\textsuperscript{10}</th>
<th>Consensus of the Three Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>% Dis</td>
<td>Total</td>
<td>% Dis</td>
</tr>
<tr>
<td>≤25</td>
<td>107</td>
<td>12</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>≤30</td>
<td>137</td>
<td>16</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>≤35</td>
<td>169</td>
<td>22</td>
<td>25</td>
<td>8</td>
</tr>
<tr>
<td>≤45</td>
<td>219</td>
<td>26</td>
<td>49</td>
<td>12</td>
</tr>
<tr>
<td>≤50</td>
<td>249</td>
<td>29</td>
<td>67</td>
<td>15</td>
</tr>
<tr>
<td>&gt;50</td>
<td>150</td>
<td>66</td>
<td>362</td>
<td>49</td>
</tr>
<tr>
<td>≥60</td>
<td>130</td>
<td>73</td>
<td>320</td>
<td>53</td>
</tr>
<tr>
<td>≥70</td>
<td>81</td>
<td>80</td>
<td>258</td>
<td>60</td>
</tr>
<tr>
<td>≥80</td>
<td>56</td>
<td>89</td>
<td>163</td>
<td>68</td>
</tr>
<tr>
<td>≥90</td>
<td>17</td>
<td>100</td>
<td>63</td>
<td>78</td>
</tr>
</tbody>
</table>

*Total=total number of patients in the training set having the probability scores indicated on the left; % Dis=percentage of the total patients with the probability scores indicated in the column to the left who had CAD.

---

\textsuperscript{11}Morise et al.\textsuperscript{11}
are sent for further testing and correctly classified, the sensitivity of the consensus approach is 94% and the specificity is 92%. The percent of correct diagnoses increased from the 67% for standard exercise ECG analysis and from the 77% for multivariable predictive equations alone to >90% correct diagnoses for the consensus approach.

Portability of the Consensus Approach

The portability of the LB-PA equation was assessed by evaluating performance characteristics across a range of disease prevalence and after varying selected variable definitions like angiographic criterion for significance.

Effect of Disease Prevalence: One hundred fifty patients were randomly chosen out of the test set to obtain seven different groups with disease prevalence ranging from 20 to 70%. Table 4 lists the percent correct diagnoses of the groups using the equations of Detrano et al., Morise et al., and LB-PA. There is very little change in the probabilities over ranges of prevalence of angiographic disease seen in clinical populations.

Effect of Angiographic Interpretation Criterion (50% vs 70% Luminal Obstruction): The angiographic criteria were changed in the test set from 70% luminal diameter narrowing to 50% narrowing. The less severe criterion caused the disease prevalence to increase from 48 to 60% (similar to the difference between the LB-PA populations and the population of Detrano et al.). Table 5 compares the percent correct diagnoses of the test set using both the 70% and 50% angiographic criteria. Comparing the first two lines, the percentages increased minimally with the 50% criteria. Was this due to the change in disease prevalence (ie., more patients will have disease with the lesser 50% criteria than with the more strict 70% criteria) or the criteria change? To resolve that question, we randomly adjusted the test set so that the disease prevalence with 50% criteria would match the disease prevalence with the 70% criteria. Test performance improved even more, suggesting that changes in angiographic criteria do not hinder the discriminatory ability of the prediction equations.

### DISCUSSION

We have demonstrated that simple clinical and exercise test variables can improve the discriminatory ability of standard exercise testing in a population of veterans referred for suspected CAD. Our derived LB-PA equation performed as well as two other validated equations in classifying patients with and without angiographically significant CAD. More importantly, application of the consensus approach to a randomly defined test set revealed characteristics that allow stratification of populations with variability in the prevalence of CAD, data element definition, and even angiographic criterion.

Variables in the Logistic Equations

The application of multivariate analysis using discriminate function and logistic regression techniques to clinical and exercise test variables have been repeatedly shown to improve on the standard application of the exercise ECG test for diagnosing CAD. A recent meta-analysis of 24 studies that considered exercise test and clinical variables to predict presence of any angiographic disease found the following variables to be significant predictors in more than half of the studies: gender, chest pain symptoms, age, elevated cholesterol level, ST slope and depression, and maximal heart rate.20 Exercise capacity, exercise-induced angina, double product, maximal systolic BP, diabetes mellitus, smoking history, abnor-

---

*Table 2—Classification of Patients Into Low, Intermediate, and High Risk for CAD in the Training Set*

<table>
<thead>
<tr>
<th>Risk Threshold</th>
<th>Palo Alto Equation</th>
<th>Equation of Detrano et al.11</th>
<th>Equation of Morise et al.10</th>
<th>Consensus of the Three Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>% Dis</td>
<td>Total</td>
<td>% Dis</td>
<td>Total</td>
</tr>
<tr>
<td>Low ≤30</td>
<td>137</td>
<td>16</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>Intermediate</td>
<td>227</td>
<td>50</td>
<td>197</td>
<td>27</td>
</tr>
<tr>
<td>High ≥75</td>
<td>65</td>
<td>82</td>
<td>213</td>
<td>64</td>
</tr>
</tbody>
</table>

*Table 2—Classification of Patients Into Low, Intermediate, and High Risk for CAD in the Training Set*

<table>
<thead>
<tr>
<th>Probability Thresholds</th>
<th>Training Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>% CAD</td>
<td>Total</td>
</tr>
<tr>
<td>Low (&lt;30)</td>
<td>70</td>
<td>4</td>
</tr>
<tr>
<td>Intermediate (30-75)</td>
<td>233</td>
<td>41</td>
</tr>
<tr>
<td>High (≥75)</td>
<td>126</td>
<td>71</td>
</tr>
</tbody>
</table>
Artificially randomly adjusting disease prevalence in the test set on the percent of correct diagnoses using the equations of Detrano et al, Morise et al, and LB-PA is shown in the first three columns. The percent of correct diagnoses is independent of the disease prevalence. In the last six columns is shown the effect of artificially randomly adjusting disease prevalence in the test set on the performance of classification according to a consensus of two of three agreement using the equations of LB-PA, Morise et al, and Detrano et al.

Among the exercise hemodynamic and ECG variables, ST segment depression, slope, exercise-induced angina, and METs have been good predictors in other studies. The logistic equation based on clinical variables outperformed standard ST criteria alone and the addition of the exercise test variables, including the ST responses, further improved the discrimination.

### Accuracy and Calibration

The equations of Morise et al and Detrano et al have been shown to be accurate in discriminating those with and without CAD in a number of populations and our new equation has now been validated in a test set. All equations performed equally well in the training and test sets despite subtle variations in the definitions of variables and weighting in the respective logistic equations; no significant differences in the ROC area or percent of correct diagnoses appeared when applying the three equations. Importantly, calibration, usually requiring an adjustment of the probability estimates according to prevalence, is not a problem with the consensus approach. Thus, the consensus approach allows for the “failure” of one equation for any reason (a patient could still be classified correctly by the other two equations) as well as for varying prevalences of CAD. Therefore, the consensus approach is portable to diverse populations with suspected CAD.

### Angiographic Criterion

There is disagreement on the criterion for clinically significant angiographic CAD. The intraobserver and interobserver variability of visual estimations of lesion severity combine with the actual underestimation of arterial atherosclerosis as demonstrated by intravascular ultrasound to make visual assessment problematic—it remains, however, the most commonly used measurement technique in clinical practice. The consensus approach may alleviate concerns about measurement accuracy by controlling for varying criteria; performance characteristics reveal excellent discriminatory ability at 50% or 70% criteria for significant luminal occlusion.

### Table 4—Effect of Artificially Randomly Adjusting Disease Prevalence in the Test Set*

<table>
<thead>
<tr>
<th>Disease Prevalence, %</th>
<th>% Correct Diagnoses</th>
<th>Classification According to Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>% Dis</td>
</tr>
<tr>
<td>20</td>
<td>64</td>
<td>3</td>
</tr>
<tr>
<td>30</td>
<td>57</td>
<td>11</td>
</tr>
<tr>
<td>40</td>
<td>55</td>
<td>11</td>
</tr>
<tr>
<td>50</td>
<td>46</td>
<td>24</td>
</tr>
<tr>
<td>60</td>
<td>38</td>
<td>29</td>
</tr>
<tr>
<td>70</td>
<td>39</td>
<td>36</td>
</tr>
</tbody>
</table>

*Effect of artificially randomly adjusting disease prevalence in the test set on the percent of correct diagnoses using the equations of Detrano et al, Morise et al, and LB-PA is shown in the first three columns. The percent of correct diagnoses is independent of the disease prevalence. In the last six columns is shown the effect of artificially randomly adjusting disease prevalence in the test set on the performance of classification according to a consensus of two of three agreement using the equations of LB-PA, Morise et al, and Detrano et al.

### Table 5—Effect of Angiographic Interpretation Criteria on the Percent of Correct Diagnoses Using the Equations of Detrano et al,¹¹ Morise et al,¹⁰ and LB-VA in the Test Set

<table>
<thead>
<tr>
<th>Angiographic Interpretation, % Narrowing</th>
<th>Disease Prevalence, %</th>
<th>Equation of Detrano et al,¹¹</th>
<th>Equation of Morise et al,¹⁰</th>
<th>LB-PA Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>48</td>
<td>76</td>
<td>76</td>
<td>77</td>
</tr>
<tr>
<td>50</td>
<td>60</td>
<td>78</td>
<td>78</td>
<td>79</td>
</tr>
<tr>
<td>50*</td>
<td>50*</td>
<td>84</td>
<td>80</td>
<td>84</td>
</tr>
</tbody>
</table>

*Artificially lowered randomly so as to match the prevalence of coronary disease using 70% narrowing criteria.
Implementation of the Consensus Approach

At our institutions, a computerized expert system (a shareware program is available from author V.F.F.) is used to generate a risk profile after calculating probabilities from the logistic equations. After the patient’s medical history, hemodynamic data, and exercise ECG data are entered into the program, the computer analyzes the data and prints out the probabilities numerically as well as the high-, intermediate-, and low-risk classification based on the consensus approach. In this way, the expert system enhances clinical care by embedding expert content into recommendations for the referring clinicians immediately after the test has been performed.

Limitations

This study is limited by clinical attributes of the patient population and by measurement issues. First, the most obvious limitation of the study is the lack of women in our study population (<2% of patients seen in the VA medical centers are female). However, the equations of Detrano et al11 and Morise et al10 have been validated in women. Second, workup bias remains a major limitation of this and other studies of test performance. Workup bias may minimize the importance of exercise capacity and maximize the importance of ST depression and angina because physicians are less likely to refer patients with a good exercise capacity for cardiac catheterization while they are more likely to refer those with markers of ischemia for catheterization. Finally, as in other studies comparing cardiac catheterization and exercise testing, this study is limited by the use of visual ECG interpretation and visual angiographic techniques. However, these techniques still predominate in clinical practice and computerized ECG analysis has not been shown to be superior to simple visual interpretation of the ST segment.23

Conclusions

We have demonstrated that by using simple clinical and exercise test variables, we can improve on the standard use of ST segment criteria to classify patients as having or not having significant CAD. Furthermore, by requiring consensus among three equations, we have made this approach portable to other populations. The consensus approach avoids the need to calibrate the equations for disease prevalence and it avoids some of the problems associated with missing data, differences in the definition of collected variables, and even in angiographic interpretation and criterion. The consensus approach is best applied utilizing a computer program to simplify the process of calculating the probability of CAD using the three equations. Application of the consensus approach can standardize the classification of patients with suspected CAD and provides a mechanism for more efficient utilization of more expensive imaging modalities, restricting their use to patients with intermediate risk. It also empowers the generalist to decide which patients need further procedures or referral to a cardiologist.

Appendix

The Study of Morise et al10

Morise et al10 studied a total of 915 consecutive patients without a history of prior myocardial infarction or coronary artery bypass surgery who were referred to the exercise laboratory at West Virginia University Medical Center between June 1981 and December 1994 for evaluation of coronary disease. All patients had coronary angiography within 3 months of the exercise test. The patients were classified as having disease if there was at least a 50% lumen diameter narrowing in one or more vessels and using this criterion, the prevalence of disease in their population was 41%. Morise et al generated both preexercise and postexercise logistic regression equations. The Morise et al preexercise test intercept and variables are as follows:

-3.6+(0.08×age)-(1.3×gender)+(0.6×symptoms)+(0.7×diabetes)+(0.3×smoking)-(1.5×body surface area)
(0.5×estrogen)+(3.0×number of risk factors)
-(0.4×rest ECG)

Gender was coded as 1 for female and 0 for male. Symptoms were classified into four categories: typical, atypical, nonanginal pain, and no pain and coded with values of 4, 3, 2, and 1, respectively. Diabetes was coded as 1 if present and 0 if absent. Smoking was coded as 2 for current smoking, 1 for any prior smoking, and 0 for never smoked. Estrogen was coded as 0 for men, 1 for estrogen negative (postmenopausal and no estrogen), and 1 for estrogen positive (premenopausal or taking estrogen). Risk factors included history of hypertension, hypercholesterolemia, and obesity (body mass index ≥27 kg/m2). Rest ECG was coded as 0 if normal and 1 if there were QRS or ST-T wave abnormalities.

Morise et al also developed a postexercise test equation for use with men and women. They found that ST depression was an independent predictor for CAD in men, but not in women. Since our study populations consist of all men, we used the postexercise test equation listed below. Along with preexercise test probability score, the postexercise equation also included ST depression, ST slope, and maximal heart rate. The equation is as follows:

\[-0.12+(4.5×pretest)+(0.37×mm ST depression)
+(1.0×ST slope)-(0.4×negative ST)
-(0.016×maximal heart rate)\]

Pretest is the pretest probability (0 to 1) derived from the pretest equation; mm ST depression was coded as 0 for women. ST slope was coded as 1 for downsloping and 0 for upsloping or horizontal. Negative ST was coded as 1 if ST depression was <1 mm depression horizontal or downsloping, or ST depression was <1.5 mm upsloping.
The Study of Detrano et al

Detrano et al included 3,549 patients from eight institutions in the United States and Europe who underwent exercise testing and angiography between 1978 and 1989. Disease was defined as >50% diameter narrowing in at least one major coronary arterial branch and the prevalence of disease according to this criterion was 64%. The selected equation components are listed below:

\[
1.9 + (0.025 \times \text{age}) - (0.6 \times \text{gender}) - (0.1 \times \text{symptoms}) - (0.05 \times \text{METs}) - (0.02 \times \text{maximal heart rate}) + (0.36 \times \text{exercise-induced angina}) + (0.6 \times \text{mm ST depression})
\]

Gender was coded as 1 for female and -1 for male. Symptoms were classified into four categories: typical, atypical, nonanginal pain, and no pain, and coded with values of 1, 2, 3, and 4, respectively. Exercise angina was coded as 1 for presence and -1 for absence.

REFERENCES